Method for the detection and/or identification of the original animal species in animal matter contained in a sample

The present invention relates to the field of the determination of an animal species, hereinafter referred to as original animal species, in a sample liable to contain an ingredient, itself obtained from at least said species. The products on which the determination according to the present invention is carried out are, for example, foods or foodstuffs intended for humans or animals, cosmetic products and, in general, products liable to contain ingredients of animal origin or, on the contrary, products in which these extracts are prohibited.

For example, identifying the animal species present in foods may be necessary in many fields of activity. A first reason is to combat fraudulent foods in which certain animal species are substituted with less expensive species, such as replacing hare with rabbit. A second reason is public health, for instance especially during the bovine spongiform encephalitis, or BSE, epidemic, a disease due to the use of animal meat meals of bovine origin for bovine feed. A third reason is religious in nature, in order to verify, for example, the absence of pork in foods. A fourth reason is legislative in nature, in particular in verifying the absence of protected species in foods.

Three main identification approaches are currently described in the literature; these methods are based on a tissue or microscopic analysis, on a protein analysis and/or on a genetic analysis.

The tissue analysis thus consists in determining the presence of bone fragments in samples of meals intended for animal feed. This technique, described in

particular in the article by Michard, Revue de l'alimentation animale [Animal feed review], vol. 508, pp 43-48, 1997, although sensitive, is laborious and is based on an expert's interpretation. It is therefore difficult to compare from one laboratory to another. In addition, by nature, it cannot detect the addition of soft tissues, such as offal, serum, blood tissues, gelatin.

Among the protein analyses used, three groups of methods for identifying animal species present in a given sample are mainly distinguished in the literature.

The first group of methods comprises protein electrophoresis techniques, which consist in detecting the soluble target proteins by specific enzymatic staining. The diagnosis is obtained after polyacrylamide gel electrophoresis, for example. However, this technique can only be carried out with fresh or frozen, unprocessed tissues, since cooking the food for a period of time is an example of processing liable to alter the proteins. This technique cannot therefore be applied to the detection of animal species present in plant meals, which undergo cooking phases during their manufacture.

The second group of methods is based on immunological techniques, using antibodies directed against soluble target proteins. The "Ouchterlony", or double immuno-diffusion, technique, a method used to differentiate antigens in a mixture, can be used. However, this technique has the major disadvantage of involving cross reactions with the epitopes of other species. The use of ELISA (enzyme-linked immunosorbent assay) techniques allows better discrimination between the species, and these techniques can be applied to cooked meat when antibodies directed against thermoresistant epitopes are used. However, problems of specificity are again

observed. By way of indication, polyclonal antibodies directed against thermoresistant epitopes from chicken are not sufficiently specific to determine whether chicken meat or turkey meat is involved.

The third group of methods comprises the chromatographic (HPLC) techniques used to characterize soluble muscle proteins. However, these techniques remain technically laborious and expensive, and can only be applied to fresh or recently frozen tissues.

The disadvantages of these three methods are mainly due to their dependence on the characterization of proteins which are thermosensitive, which denature when the foods are cooked for a period of time and which lose their biological activity after the animal's death, and the presence of which often depends on the cell type that is examined.

It is thus preferable to directly analyze the DNA, rather than the proteins, of the sample, in order to identify the original animal species which is or are present in a given sample, the DNA being identical in all the cell types of the same animal and stable by comparison with the proteins. A third approach therefore consists in analyzing the DNA present in the sample. Only recently have methods based in particular on the use of restriction enzymes or of genetic markers thus been found in the literature, these methods having the advantage of being able to be applied to processed products, in particular after thermal treatment.

The nucleic acid determination may make use of restriction enzymes, or the technique referred to as RFLP (Restriction Fragment Length Polymorphism, see in particular Meyer et al., Journal of AOAC International, vol 78 No. 6, pp 1542-1551, 1995). The restriction enzymes cleave the DNA, extracted beforehand from the sample to be analyzed, at precise sites in the

macromolecule. It then suffices to compare, by simple electrophoresis, the fragments obtained with those of control samples representative of the species to be identified. However, the analysis of the results obtained by this technique is very tricky, in particular when several animal species are present in the sample.

The nucleic acid determination can also consist in sequencing a ubiquitous marker, such as mitochondrial DNA cytochrome B. Mitochondrial DNA is a known target for this type of analysis since each mitochondrion contains from one to ten mitochondrial DNA molecules, and each cell contains from a few tens to a few thousand mitochondria, which makes it possible to work on a very small amount of sample. Thus, Bartlett & Davidson (Biotechniques, vol. 12, No. 3, 1992) describe a method called FINS (Forensically Informative Nucleotide Sequencing). This method consists in i) isolating the DNA present in a biological sample, ii) amplifying this DNA by PCR using primers specific for the mitochondrial cytochrome B gene, the primers being chosen in the portion of the gene which is highly conserved during evolution, and iii) sequencing the amplified DNA segment. The sequence is then used for a phylogenetic analysis by means of a database, allowing identification of the animal species initially present in the sample. While this method has the advantage of being rapid and usable on any type of foods (fresh, frozen, processed, etc.), it nevertheless has the major disadvantage of not enabling the analysis of mixtures of species, from mixtures of amplified sequences derived from the same ubiquitous polymorphic marker, and thus remains reserved for homogeneous starting materials.

The analysis can also consist in amplifying a marker specific for a given species. Thus, Lahiff et al. (Molecular and Cellular Probes, vol. 15, pp 27-35,

2001) describe the identification of an ovine, bovine or avian species present in a sample using, by PCR, particular primers specific to each species. A method developed by S. Colgan et al. was also described in 2001 (FOOD Research International, 2001, vol 34, No. 5, 401-414), for detecting 4 species in a mixture using specific primers by PCR. While this method makes it possible to specifically and rapidly identify such and such a species, it cannot be applied simultaneously to the detection of several species. Successive PCRs are then necessary if the detection of several species is desired. The detection of six animal species using a multiplex PCR (*Matsunaga* et al. 1999 Meat Sciences, (1999), 145-148) and (Matsunaga T., *et al., Nippon Shokuhin KogakuKaishi, (1999) vol 46. No. 3, 187-194)* is thus found in the prior art. However, this technique remains tricky and difficult to apply and, in practice, involves prior knowledge of the species sought. This technique cannot, however, be applied blind, i.e. without prior knowledge of the species likely to be present in the sample. It does not make it possible to have quantitative results because of the difficulties due to the multiplex amplification and the possibilities of mismatches. In addition, this technique requires a large number of specific primers if the intention is to test a large number of species, which is relatively impossible to realize in practice due to problems of sensitivity and specificity. Finally, if a species is not represented in the set of primers but is nevertheless present in the sample to be analyzed, the result will be distorted.

The techniques described above make it possible to determine, without prior knowledge, the species present when the sample comprises only one species, and they make it possible to detect several species when there is prior knowledge of the species brought together, but none of the techniques mentioned above allows a determination in the presence of a mixture of several

species without prior knowledge of said species brought together. In addition, most of the techniques described above, when several species are present, do not allow a reliable determination when the proportions of the various species are very different in the sample.

There is therefore a great need for a technique which, while remaining generic, can detect one or more species, even present in large number in the same sample to be analyzed or in very small amount, and without prior knowledge of the species present.

In fact, while, in a product, the unwanted species must be present in amounts greater than 5% or even 1% according to the legislation, relative to the species normally present in order for there to be fraudulent practice, which eases the required performance levels for the molecular diagnostic test, it is quite different in the case of products in which the presence of products of animal origin is prohibited. For example, in the case of meals used in France for animal feed since January 1st, 2001, traces of content of product of animal origin are sought, and the technical constraint is considerable in terms of sensitivity of the method since most of the material is of plant origin and the addition of animal material ranges between 0.1 and 5% weight/weight.

A need therefore exists for a determining tool which allows the qualitative and/or quantitative identification or detection of animal species, blind, i.e. without a priori regarding the identity of the species sought, which can be used simply, while remaining specific, reliable and accurate, and which can be used in a medium possibly containing ingredients obtained from several animal species.

The problem to be solved is of considerable complexity. The determination must be possible blind, i.e. the

sample may or may not contain ingredients obtained from one or more animal species and these original species are unknown. If the sample contains ingredients obtained from animal species, the original species must be determined and may be related, and it must be possible to make the determination by carrying out just one analysis, with a single reagent and a single amplification step, without a prior step for predetermining, for example, the group of species or without using batteries of tests making it possible, for example, to classify the reagents by genera or species so as to avoid, for example, cross reactions.

To this effect, the applicant has discovered a set of sequences consisting of the group comprising the sequences SEQ ID Nos 1 to 232, 242 to 261, the sequences respectively complementary thereto, and any homologous sequences, comprising at least 5 contiguous monomers included in any one of said sequences and exhibiting at least 70% identity with said any sequence, which make it possible, using "molecular biology" analytical methods, to determine at least one original animal species in a sample liable to contain an ingredient obtained from at least said species.

Before disclosing the invention, various terms used in the description and the claims are defined hereinafter.

- A "determination" is understood to be the identification or the quantitative and/or qualitative detection or analysis of an animal species.

- An "animal species" is understood to be the simplest category used in the classification of living species or taxonomy. Living species are classified in categories called taxa; the most important taxa are the kingdom (plant or animal), the phyllum or division, the class, the order, the family, the genus and the species. Birds, fish and mammals are classes of

vertebrate animals.

- The term "original animal species" is understood to mean the animal species of the animal whose tissues, whatever they are, were used as starting material for preparing the ingredient(s) of the sample of the product subjected to the determination according to the present invention.

- A "molecular biology method" is a method based on the enzymatic amplification of nucleic acid (DNA and/or RNA) targets in vitro and the use of oligonucleotide probes.

- A "sample" is any part obtained directly or indirectly from a starting product, matter or material, itself liable to contain at least one ingredient obtained from at least one "original" animal species. As a consequence of this definition, the sample to be determined in accordance with the present invention is liable to contain said ingredient of animal origin, based on which the animal species which has or have made up or constituted the starting product, matter or material is or are identified or detected. For the purpose of the present invention, the starting product can be a biological material, a food or foodstuff, for example based on meat or fish, a cosmetic product, etc.

- The term "lysis step" is understood to mean a step capable of releasing the nucleic acids contained in the protein and/or lipid envelopes of the microorganisms (such as cell debris which disturbs the subsequent reactions). By way of example, use may be made of the lysis methods as described in the applicant's patent applications:

WO-A-00/05338 regarding mixed magnetic and mechanical lysis,
WO-A-99/53304 regarding electrical lysis, and

WO-A-99/15321 regarding mechanical lysis.

Those skilled in the art may use other well-known lysis methods, such as thermal or osmotic shocks or chemical lyses with chaotropic agents such as guanidium salts (US-A-5,234,809).

- The term "purification" is understood to mean separation between the nucleic acids and the other cell components released in the lysis step. This step generally makes it possible to concentrate the nucleic acids. By way of example, it is possible to use magnetic particles optionally coated with oligonucleotides, by adsorption or covalence (on this subject, see patents US-A-4,672,040 and US-A-5,750,338), and thus to purify the nucleic acids which are attached to these magnetic particles, by means of a washing step. This nucleic acid purification step is particularly advantageous if it is desired to subsequently amplify said nucleic acids. A particularly advantageous embodiment of these magnetic particles is described in the patent applications filed by the applicant under the following references: WO-A-97/45202 and WO-A-99/35500.

In the latter of these patent applications, the particles are thermosensitive magnetic particles which each have a magnetic core covered with an intermediate layer. The intermediate layer is itself covered with an outer layer based on a polymer capable of interacting with at least one biological molecule, for example nucleic acid; the outer polymer is thermosensitive and has a predetermined lower critical solution temperature (LCST) of between 10 and 100°C, and preferably between 20 and 60°C. This outer layer is synthesized from cationic monomers which generate a polymer having the ability to bind nucleic acids. This intermediate layer isolates the core's magnetic forces in order to avoid problems of inhibition of the techniques for amplifying

these nucleic acids.

Another advantageous example of a method for purifying nucleic acids is the use of silica, either in the form of a column (Qiagen kits, for example), or in the form of inert particles [Boom R. et al., J. Clin. Microbiol., 1990, No. 28(3), p. 495-503] or magnetic particles (Merck: MagPrep® Silica, Promega: MagneSil™ Paramagnetic particles). Other very widely used methods are based on ion exchange resins in a column (Qiagen kits, for example) or in a paramagnetic particulate format (Whatman: DEAE-Magarose) [Levison PR et al., J. Chromatography, 1998, p. 337-344]. Another method which is very relevant but not exclusive for the invention is that of adsorption onto a metal oxide support (Xtrana: Xtra-Bind™ matrix).

- A "sequence", or a "nucleotide fragment", or an oligonucleotide or a polynucleotide, is a chain of nucleotide units assembled together via phosphoester bonds, characterized by the informational sequence of the natural nucleic acids capable of hybridizing with a nucleotide fragment, it being possible for the chain to contain monomers having different structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis.

- A "unit" is derived from a monomer which may be a natural nucleotide of nucleic acid, of which the constituent elements are a sugar, a phosphate group and a nitrogenous base; in DNA, the sugar is 2-deoxyribose, and in RNA, the sugar is ribose; depending on whether it is a question of DNA or RNA, the nitrogenous base is chosen from adenine, guanine, uracil, cytosine and thymine; or alternatively the monomer is a nucleotide which has been modified in at least one of the three constituent elements; by way of example, the modification can affect either the bases, with modified bases such as inosine, 5-methyldeoxycytidine,

deoxyuridine, 5-dimethylaminodeoxyuridine, 2,6-diaminopurine, 5-bromodeoxyuridine or any other modified base capable of hydridization, or the sugar, for example the replacement of at least one deoxyribose with a polyamide (P.E. Nielsen et al., Science, 254, 1497-1500 (1991)), or alternatively the phosphate group, for example replacement thereof with esters chosen in particular from diphosphates, alkyl- and arylphosphonates and phosphorothioates.

- The term "informational sequence" is understood to mean any ordered series of units of nucleotide type, the chemical nature of which and the order of which in a reference direction constitute an item of information of the same quality as that of the natural nucleic acids.

- The term "hybridization" is understood to mean the process during which, under suitable conditions, two nucleotide fragments having sufficiently complementary sequences are capable of forming a double strand with stable and specific hydrogen bonds. A nucleotide fragment "capable of hybridizing" with a polynucleotide is a fragment which can hybridize with said polynucleotide under hybridization conditions which can be determined in a known manner in each case. The hybridization conditions are determined by means of the stringency, i.e. the severity of the operating conditions. The higher the stringency at which the hybridization is carried out, the more specific the hybridization is. The stringency is defined in particular according to the base composition of a probe/target duplex, and also by means of the degree of mismatching between two nucleic acids.

The "stringency" can also depend on the parameters of the reaction, such as the concentration and the type of ion species present in the hybridization solution, the nature and the concentration of denaturing agents

and/or the hybridization temperature. The stringency of the conditions under which a hybridization reaction should be carried out will depend mainly on the target probes used. All these data are well known and the appropriate conditions can be determined by those skilled in the art.

In general, depending on the length of the target probes used, the temperature for the hybridization reaction is between approximately 20 and 70°C, in particular between 35 and 65°C, in a saline solution at a concentration of approximately 0.5 to 1 M.

- A "probe" comprises a nucleotide fragment comprising from 5 to 100 monomers, in particular from 6 to 35 monomers, possessing a hybridization specificity under given conditions so as to form a hybridization complex with a nucleotide fragment having, in the present case, a nucleotide sequence included, for example, in a ribosomal RNA, the DNA obtained by reverse transcription of said ribosomal RNA, and the DNA (referred to here as ribosomal DNA or rDNA) for which said ribosomal RNA is the transcription product; a probe can be a capture probe or a detection probe.

- A "capture probe" is immobilized or can be immobilized on a solid support by any suitable means, i.e. directly or indirectly, for example by covalence or adsorption.

- A "detection probe" can be labeled by means of a label chosen from radioactive isotopes, enzymes (in particular a peroxidase, an alkaline phosphatase, or an enzyme capable of hydrolyzing a chromogenic, fluorigenic or luminescent substrate), chemical chromophore compounds, chromogenic, fluorigenic or luminescent compounds, nucleotide base analogs, and ligands such as biotin.

- A "primer" comprises a nucleotide fragment comprising from 5 to 100 nucleotide units and possessing a hybridization specificity under given conditions for the initiation of an enzymatic polymerization, for example in an amplification technique, in a sequencing process, in a reverse transcription method, etc.

- "The homology" characterizes the degree of identity of two compared nucleotide fragments, for which the criteria selected for the present invention are defined below.

The probes and primers according to the invention are chosen from:

(a) the sequences identified in the sequence listing attached in the appendix to the description,

(b) the sequences complementary to each of the sequences identified in the sequence listing attached in the appendix to the description, the complementarity meaning any sequence capable of hybridizing, at a temperature of between 20 and 70°C, and preferably between 35 and 65°C, in saline solution at a concentration of approximately 0.5 to 1 M, and preferably 0.8 to 1 M, with any one of the sequences identified in the sequence listing attached in the appendix to the description,

(c) the sequences homologous to each of the sequences identified in the sequence listing attached in the appendix to the description, and of the sequences complementary to each of the sequences identified in the sequence listing attached in the appendix to the description, respectively, the homology meaning any sequence, for example fragment, comprising a series of at least 5 contiguous nucleotides included in any one of said sequences, and exhibiting at least 70% identity with said any sequence; by way of example, a fragment (c) comprises 10 nucleotides, among which 5 contiguous nucleotides belong to a sequence (a) and at least two

nucleotides of the remaining 5 nucleotides are identical, respectively, to the two corresponding nucleotides in the reference sequence, after alignment.

- The term "identifying sequence" denotes any sequence or any fragment as defined above, which can serve as a detection probe and/or capture probe.

- The term "detection" is understood to mean either a direct detection by means of a physical method, or a method of detection using a label.

Many detection methods exist for detecting nucleic acids [see, for example, Kricka et al., Clinical Chemistry, 1999, No. 45(4), p. 453—458 or Keller G.H. et al., DNA Probes, 2nd Ed., Stockton Press, 1993, sections 5 and 6, p. 173-249].

The term "label" is understood to mean a tracer capable of engendering a signal. A nonlimiting list of these tracers comprises the enzymes which produce a signal that can be detected, for example, by colorimetry, fluorescence or luminescence, such as horseradish peroxidase, alkaline phosphatase, beta-galactosidase, glucose-6-phosphate dehydrogenase; chromophors such as fluorescent, luminescent or dye compounds; electron dense groups which can be detected by electron microscopy or by means of their electrical properties such as conductivity, by amperometry or voltametry methods, or by impedance measurements; groups which can be detected by optical methods such as diffraction, surface plasmon resonance or contact angle variation or by physical methods such as atomic force spectroscopy, tunnel effect, etc.; radioactive molecules such as ^{32}P, ^{35}S or ^{125}I.

Thus, the polynucleotide can be labeled during the enzymatic amplification step, for example by using a labeled triphosphate nucleotide for the amplification

reaction. The labeled nucleotide will be a deoxyribonucleotide in amplification systems generating a DNA, such as PCR, or a ribonucleotide in amplification techniques generating an RNA, such as the TMA or NASBA techniques.

The polynucleotide can also be labeled after the amplification step, for example by hybridizing a labeled probe according to the sandwich hybridization technique described in document WO-A-91/19812.

Another particularly preferred method for labeling nucleic acids is described in the applicant's application FR-A-2 780 059. Another preferred method of detection uses the 5'-3' exonuclease activity of a polymerase, as described by Holland P.M., PNAS (1991) p 7276-7280.

Signal amplification systems can be used as described in document WO-A-95/08000 and, in this case, the preliminary enzymatic amplification reaction may not be necessary.

- The term "enzymatic amplification" is understood to mean a process generating multiple copies of a particular nucleotide fragment using specific primers by means of the action of at least one enzyme. Thus, for nucleic acid amplification, there exists, inter alia, the following techniques:

  - PCR (Polymerase Chain Reaction), as described in patents US-A-4,683,195, US-A-4,683,202 and US-A-4,800,159,
  - LCR (Ligase Chain Reaction), disclosed, for example, in patent application EP-A-0 201 184,
  - RCR (Repair Chain Reaction), described in patent application WO-A-90/01069,
  - 3SR (Self Sustained Sequence Replication) with patent application WO-A-90/06995,

- NASBA (Nucleic Acid Sequence-Based Amplification) with patent application WO-A-91/02818, and
- TMA (Transcription Mediated Amplification) with patent US-A-5,399,491.

The term "amplicons" is then used to denote the polynucleotides generated by means of an enzymatic amplification technique.

- The term "solid support" as used here includes all the materials on which a nucleic acid can be immobilized. Synthetic materials or natural materials, optionally chemically modified, can be used as a solid support, in particular polysaccharides such as cellulose-based materials, for example paper, cellulose derivatives such as cellulose acetate and nitrocellulose or dextran, polymers, copolymers, in particular based on monomers of the styrene type, natural fibers such as cotton, and synthetic fibers such as nylon; inorganic materials such as silica, quartz, glasses, ceramics; latices; magnetic particles; metal derivatives, gels, etc. The solid support can be in the form of a microtitration plate, of a membrane as described in application WO-A-94/12670, of a particle or of a biochip.

- The term "biochip" is understood to mean a solid support which is small in size and to which is attached a multitude of capture probes at predetermined positions.

By way of illustration, examples of these biochips are given in the publications by [G. Ramsay, Nature Biotechnology, 1998, No. 16, p. 40-44; F. Ginot, Human Mutation, 1997, No. 10, p. 1-10; J. Cheng et al., Molecular diagnosis, 1996, No. 1(3), p. 183-200; T. Livache et al., Nucleic Acids Research, 1994, No. 22(15), p. 2915-2921; J. Cheng et al., Nature Biotechnology, 1998, No. 16, p. 541-546] on in patents

US-A-4,981,783, US-A-5,700,637, US-A-5,445,934, US-A-5,744,305 and US-A-5,807,522.

The main characteristic of the solid support should be to conserve the characteristics of hybridization of the capture probes to the nucleic acids while at the same time generating a minimum background noise for the detection method. An advantage of biochips is that they simplify the use of many capture probes, thus allowing multiple detection of the species to be detected.

The invention described hereinafter makes it possible to solve the problems posed by the methods described above, equally in terms of sensitivity, specificity, multidetection capacity and identification, while at the same time being rapid and easy to implement.

The invention relates to a method for determining an original animal species in a sample liable to contain an ingredient obtained from at least said species, characterized in that:

a) a nucleic acid fraction obtained from said sample is provided,

b) at least one reagent specific for the animal species is provided, chosen from the group consisting of

- the reference sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,
- the sequences complementary to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, respectively, the complementarity meaning any sequence capable of hybridizing, at a temperature of between 20 and 70°C, and preferably between 35 and 65°C, in saline solution at a concentration of approximately 0.5 to 1 M, and preferably 0.8 to 1 M, with any one of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,
- the sequences homologous to each of the

sequences SEQ ID Nos 1 to 232, and Nos 242 to 261 and of the sequences complementary to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, respectively, the homology meaning any sequence, for example fragment, comprising a series of at least 5 contiguous nucleotides included in any one of said sequences, and exhibiting at least 70% identity with said any sequence,

c) the nucleic acid fraction and said reagent are brought into contact, and

d) any signal or item of information resulting from the specific reaction between said reagent and the nucleic acid fraction, characterizing the presence in said sample of said original animal species, is determined by means of detection.

It also relates to a method as described above, characterized in that a set comprising a multiplicity of said reagents specific for the same original species and/or for respectively different original animal species is provided; and a multiplicity of signals or items of information characterizing the presence in said sample of the same original animal species or of several respectively different original animal species is determined.

It also relates to any nucleotide sequence characterized in that it is chosen from the group consisting of:

a) the reference sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,

b) the sequences complementary to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, respectively, the complementarity meaning any sequence capable of hybridizing, at a temperature of between 20 and 70°C, and preferably between 35 and 65°C, in saline solution at a concentration of approximately 0.5 to 1 M, and preferably 0.8 to 1 M, with any one of the

sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,

c) the sequences homologous to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, and of the sequences according to b), respectively, the homology meaning any sequence, for example fragment, comprising a series of at least 5 contiguous nucleotides included in any one of said sequences, and exhibiting at least 70% identity with said any sequence.

It also relates to the use of a sequence defined above, for determining at least one original animal species in a sample liable to contain an ingredient obtained from at least said animal species.

The invention relates to a method for determining an original animal species in a sample liable to contain an ingredient obtained from at least said species, characterized in that it allows said determination in a sample containing at least one other ingredient obtained from another animal species and without prior knowledge of the species brought together, and in that:

a) a nucleic acid fraction obtained from said sample is provided,

b) at least one reagent specific for the animal species is provided, chosen from the group consisting of:

- the reference sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,

- the sequences complementary to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, respectively, the complementarity meaning any sequence capable of hybridizing, at a temperature of between 20 and 70°C, and preferably between 35 and 65°C, in saline solution at a concentration of approximately 0.5 to 1 M, and preferably 0.8 to 1 M, with any

one of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261,

- the sequences homologous to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, and of the sequences complementary to each of the sequences SEQ ID Nos 1 to 232, and Nos 242 to 261, respectively, the homology meaning any sequence, for example fragment, comprising a series of at least 5 contiguous nucleotides included in any one of said sequences, and exhibiting at least 70% identity with said any sequence,

c) the nucleic acid fraction and said reagent are brought into contact, and

d) any signal or item of information resulting from the specific reaction between said reagent and the nucleic acid fraction, characterizing the presence in said sample of said original animal species, is determined by means of detection.

The invention can also be a probe for determining at least one original animal species, comprising at least one identifying nucleotide sequence defined above.

It also relates to a primer for the specific amplification of a nucleic acid from an original animal species, comprising at least one identifying nucleotide sequence defined above.

Another embodiment of the invention is a reagent for determining at least one original animal species, comprising a solid support, which may or may not be divided up, to which a nucleotide sequence defined above is attached.

According to the invention, the nucleotide sequences or

their fragments can be attached to a solid support and can constitute a biochip which makes it possible to determine the multiplicity of signals or items of information.

The method according to the invention can be carried out manually, semi-automatically or automatically, allowing the use of a means for determining the original animal species in animal matter contained in a sample.

This invention also relates to a method of detection using in particular the biochip technique. This method of detection is specific for the species being sought by virtue of the use of sequences, referred to as identifying sequences for each species, as a probe. The rapidity, the sensitivity and the specificity of this method of detection make it possible to apply it equally to any medium. In particular, this method applies to any sample of a food product comprising animal matter, whatever its condition and the methods of manufacture and/or of production used, in particular the cooking, dehydration and/or storage techniques, and to any sample of a manufactured product liable to contain animal extracts, such as, for example, cosmetic products and/or pharmaceutical products comprising, for example, gelatins of animal origin.
This simultaneous single-step detection of multiple specific amplification products is possible by virtue of the use of a solid support, in particular in the form of a solid support which is small in size and to which is attached a multitude of capture probes at predetermined positions, or "biochip", these capture probes consisting of a set of fragments of, or of all, nucleotide sequences specific for said identifying sequences for the species being sought.

These nucleotide sequences can also be used in all the known hybridization techniques, such as the "Dot-blot"

techniques for depositing a spot onto a filter [Maniatis et al., Molecular Cloning, Cold Spring Harbor, 1982], the "Southern blot" techniques for transferring DNA [Southern E.M., J. Mol. Biol., 1975, 98, 503], the "Northern blot" techniques for transferring RNA, or the "Sandwich" techniques [Dunn A.R. et al., Cell, 1977, 12,23].

The present invention also relates to the determination of a group of species or class of animal species or taxon. These groups of species or classes or taxa consist, for example, of a class, such as the class of mammals, birds or fish, or even of subgroups of species such as a family of birds or of two subgroups combined, such as birds or mammals.

This identification is possible through the identification of nucleotide sequences, called signature sequences, characteristic of a class, of a group, of a subgroup or of a taxon, and corresponding to regions which have been conserved for all the individuals making up the group. Any signature sequence specific for a class of animals, used in the method according to the present invention, exhibits the characteristic according to which, firstly, it has a nucleic acid region which has been conserved for virtually all the animal species of the same taxonomic class and, secondly, it can be distinguished from other sequences corresponding to the same definition as above, under the usual conditions for determination, defined generically in the attached claims.

The invention also relates to a method for determining a group of original animal species in a sample liable to contain an ingredient obtained from at least one species belonging to said group of animal species under consideration, characterized in that:

a) a nucleic acid fraction obtained from said sample is provided,

b) the nucleotide sequence(s) characteristic of the group of animal species to be determined is (are) identified,

c) at least one reagent comprising a sequence identified in step b) is provided,

c) the nucleic acid fraction and said reagent are brought into contact, and

d) any signal or item of information resulting from the presence of one of the sequences defined above, characterizing the presence in said sample of a group of original animal species, is determined by means of detection.

For example, for detecting the presence of mammals, use will be made of:

1/ the signature sequence M1, corresponding to the sequence SEQ ID No. 235 GACA**CAA**CAA CAGC, positions 14685 to 14698 (genbank *Bos taurus* reference sequence; accession No. V00654). The CAA bases at positions 14689-14690-14691 (genbank *Bos taurus* reference sequence; accession No. V00654) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of these three bases at the positions indicated above thus makes it possible to determine the presence of mammals in the sample;

2/ the signature sequence M2, corresponding to the sequence SEQ ID No. 262, positions 14634 to 14648 (genbank *Bos taurus* reference sequence; accession No. V00654). The T base at position 14641 (genbank *Bos taurus* reference sequence; accession No. V00654) is conserved for all the nucleic acid material corresponding to the predefined species making up the

group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of this base at the position indicated above thus makes it possible to determine the presence of mammals in the sample;

3/ the signature sequence M3, corresponding to the sequence SEQ ID No. 263, positions 14771 to 14785 (genbank *Bos taurus* reference sequence; accession No. V00654). The A base at position 14778 (genbank *Bos taurus* reference sequence; accession No. V00654) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of this base at the position indicated above thus makes it possible to determine the presence of mammals in the sample.

Identification of the presence of birds is determined by means of the signatures:

1/ O1, corresponding to the sequence SEQ ID No. 236 TCC**CT**AGCCT TCTC, positions 15073 to 15086 (*Gallus gallus* reference sequence; genbank accession No. X52392). The **CT** bases (positions 15076-15077) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of these two bases at the positions indicated above thus makes it possible to determine the presence of birds in the sample.

2/ O2, corresponding to the sequence SEQ ID No. 237 ACA**CT**TGCCG GAAC, positions 15098 to 15111 (*Gallus gallus* reference sequence; genbank accession No. X52392). The **CT** or **CA** bases (positions 15101-15102) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of these two bases at the positions indicated above thus makes it possible to determine the presence of birds in the sample.

3/ O3, corresponding to the sequence SEQ ID No. 264, positions 15036 to 15050 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15043 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

4/ O4, corresponding to the sequence SEQ ID No. 265, positions 15069 to 15083 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15076 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up

the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

5/ O5, corresponding to the sequence SEQ ID No. 266, positions 15094 to 15108 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15101 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

6/ O6, corresponding to the sequence SEQ ID No. 267, positions 15102 to 15116 (genbank *Gallus gallus* reference sequence; accession No. X52392). The A base at position 15109 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

7/ O7, corresponding to the sequence SEQ ID No. 268, positions 15108 to 15122 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15115 (genbank *Gallus gallus* reference

sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

8/ O8, corresponding to the sequence SEQ ID No. 269, positions 15232 to 15246 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15239 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

Identification of the presence of mammals and of birds is determined by means of the signature V, corresponding to the sequence SEQ ID No. 238 ATAGCCACAGCATT, positions 14883 to 14896 (genbank *Bos taurus* reference sequence; accession No. V00654). The **GC** bases (at positions 14886 and 14887) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds and mammals. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds and mammals. The presence of these two

bases at the positions indicated above thus makes it possible to determine the presence of mammals and of birds in the sample.

Identification of the presence of fish is determined by means of:

1/ the signature P1, corresponding to the sequence SEQ ID No. 239 **ATAATA**ACCTCTTT, positions 14713 to 14726 (*Gadus morhua* reference sequence; genbank accession No. X99772). The **ATA** or **ATG** bases (positions 14716-14717-14718) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case fish. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of these three bases at the positions indicated above thus makes it possible to determine the presence of fish in the sample;

2/ the signature sequence of P2, corresponding to the sequence SEQ ID No. 270, positions 14512 to 14526 (genbank *Gadus morhua* reference sequence; accession No. X99772). The T base at position 14519 (genbank *Gadus morhua* reference sequence; accession No. X99772) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case fish. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of this base at the position indicated above thus makes it possible to determine the presence of fish in the sample;

3/ the signature sequence P3, corresponding to the sequence SEQ ID No. 271, positions 14710 to 14724 (genbank *Gadus morhua* reference sequence; accession No.

X99772). The T base at position 14717 (genbank *Gadus morhua* reference sequence; accession No. X99772) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case fish. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of this base at the position indicated above thus makes it possible to determine the presence of fish in the sample.

The present invention therefore also relates to a nucleotide sequence, characterized in that it is chosen from the group consisting of:

a) the reference sequences SEQ ID Nos 235 to 239, and 262 to 271,

b) the sequences complementary to each of the sequences SEQ ID Nos 235 to 239, and 262 to 271, respectively, the complementarity meaning any sequence capable of hybridizing, at a temperature of between 20 and 70°C, in saline solution at a concentration of approximately 0.5 to 1M, with any one of the sequences SEQ ID Nos 235 to 239, and 262 to 271,

c) the sequences homologous to each of the sequences SEQ ID Nos 235 to 239, and 262 to 271, and of the sequences according to b), respectively, the homology meaning any sequence, for example fragment, comprising a series of at least 5 contiguous nucleotides included in any one of said sequences and also a group of two or three nucleotides belonging to a region which has been conserved for all the species of a group under consideration, and said sequence exhibiting at least 70% identity with said any sequence.

It relates more particularly to the nucleotide sequences as defined above, and characterized in that they consist of a group of 2 to 3 nucleotides included

in one of the sequences SEQ ID Nos 235 to 239 and corresponding to a region which has been conserved for all the species of a group under consideration.

It also relates to the use of the sequences defined above, that is to say characterized in that they consist of a group of 2 to 3 nucleotides included in one of the sequences SEQ ID Nos 235 to 239 and corresponding to a region which has been conserved for all the species of a group under consideration, for determining a group of original animal species in a sample liable to contain an ingredient obtained from at least one animal species belonging to said group of animal species under consideration.

These sequences, termed signature sequences, are chosen from the group consisting of the nucleotide sequence consisting of the CAA bases at positions 14689-14690-14691 of SEQ ID No. 235, the nucleotide sequence consisting of the CT bases at positions 15076-15077 of SEQ ID No. 236, the nucleotide sequence consisting of the CT bases at positions 15101-15102 of SEQ ID No. 237, the nucleotide sequence consisting of the GC bases at positions 14886-14887 of SEQ ID No. 238, and the nucleotide sequence consisting of the ATA bases at positions 14713-14726 of SEQ ID No. 239.

It relates more particularly to the nucleotide sequences as defined above, and characterized in that they consist of 1 nucleotide included in one of the sequences SEQ ID Nos 262 to 271 and corresponding to a region which has been conserved for all the species of a group under consideration.

It also relates to the use of the sequences defined above, that is to say characterized in that they consist of one nucleotide included in one of the sequences SEQ ID Nos 262 to 271 and corresponding to a region which has been conserved for all the species of

a group under consideration, for determining a group of original animal species in a sample liable to contain an ingredient obtained from at least one animal species belonging to said group of animal species under consideration.

These sequences, termed signature sequences, are chosen from the group consisting of the nucleotide sequence consisting of the T base at position 14641 of SEQ ID No. 262, the nucleotide sequence consisting of the A base at position 14778 of SEQ ID No. 263, the nucleotide sequence consisting of the C base at position 15043 of SEQ ID No. 264, the nucleotide sequence consisting of the C base at position 15076 of SEQ ID No. 265, the nucleotide sequence consisting of the C base at position 15101 of SEQ ID No. 266, the nucleotide sequence consisting of the A base at position 15109 of SEQ ID No. 267, the nucleotide sequence consisting of the C base at position 15115 of SEQ ID No. 268, the nucleotide sequence consisting of the C base at position 15239 of SEQ ID No. 269, the nucleotide sequence consisting of the T base at position 14519 of SEQ ID No. 270, and the nucleotide sequence consisting of the T base at position 14717 of SEQ ID No. 271.

It also relates to a reagent for determining at least one original animal species, comprising a solid support, which may or may not be divided up, to which a nucleotide sequence chosen from the group consisting of the sequences SEQ ID Nos 235 to 239, and Nos 262 to 271, is attached.

It also relates to the method for determining a group of original animal species in a sample liable to contain an ingredient obtained from at least one species belonging to said group of animal species under consideration, characterized in that:

a) a nucleic acid fraction obtained from

said sample is provided,

b) at least one reagent comprising a sequence defined above is provided,

c) the nucleic acid fraction and said reagent are brought into contact, and

d) any signal or item of information resulting from the presence of one of the signature sequences chosen from the group consisting of the nucleotide sequence consisting of the CAA bases at positions 14689-14690-14691 of SEQ ID No. 235, the nucleotide sequence consisting of the CT bases at positions 15076-15077 of SEQ ID No. 236, the nucleotide sequence consisting of the CT bases at positions 15101-15102 of SEQ ID No. 237, the nucleotide sequence consisting of the GC bases at positions 14886-14887 of SEQ ID No. 238, and the nucleotide sequence consisting of the ATA or ATG bases at positions 14713-14726 of SEQ ID No. 239, the nucleotide sequence consisting of the T base at position 14641 of SEQ ID No. 262, the nucleotide sequence consisting of the A base at position 14778 of SEQ ID No. 263, the nucleotide sequence consisting of the C base at position 15043 of SEQ ID No. 264, the nucleotide sequence consisting of the C base at position 15076 of SEQ ID No. 265, the nucleotide sequence consisting of the C base at position 15101 of SEQ ID No. 266, the nucleotide sequence consisting of the A base at position 15109 of SEQ ID No. 267, the nucleotide sequence consisting of the C base at position 15115 of SEQ ID No. 268, the nucleotide sequence consisting of the C base at position 15239 of SEQ ID No. 269, the nucleotide sequence consisting of the T base at position 14519 of SEQ ID No. 270, and the nucleotide sequence consisting of the T base at position 14717 of SEQ ID No. 271,

characterizing the presence in said sample of a class of original animal species or of a group of original animal species, is determined by means of detection.

The identifying sequences can also be used as specific primers in PCR identification techniques, by mixing several primers chosen from the nucleotide sequences specific for an animal species in the presence of other species liable to be present in the media to be assayed, and in that at least one of said primers is chosen from the group consisting of the sequences SEQ ID Nos 1 to 232, and 242 to 261, and any sequences comprising at least 5 contiguous monomers included in any one of said sequences and exhibiting at least 70% identity with said any sequence.

The invention also relates to the nucleotide sequences chosen from the group consisting of the sequences SEQ ID No. 240 to SEQ ID No. 241 and SEQ ID Nos 272 to 276, and to their use as universal amplification primers, that is to say primers which can be used for detecting species in a mixture and which are sufficiently sensitive, with respect to various species, to avoid erroneous results due to the masking of certain species present in a very small proportion, because of too great a sensitivity with respect to another species liable to be present in a larger proportion. These primers are preferably used as pairs chosen from the following pairs: SEQ ID No. 240 and SEQ ID No. 241, SEQ ID No. 272 and SEQ ID No. 273, and SEQ ID No. 274 and SEQ ID No. 275.

These primers are used for carrying out the amplification steps of the methods described above, in particular when the samples comprise or are liable to contain biological material originating from species belonging to the vertebrate group.

The following examples are given by way of illustration and are in no way limiting in nature. They will make it possible to understand the invention more fully.

Example 1: Detection of an animal species in a sample (table 1)

a) Preparation of the sample

Samples originating from several animal species (mammals, birds, fish) were used in this example. The samples could be divided up into several categories:

- reference samples (denoted "ref" in table 1):

- reference DNA from various animal species: mammalian DNA (cattle, goat, sheep, pig, rabbit, hare, reindeer), bird DNA (ostrich, chicken, turkey, goose), fish DNA (cod, yellowfin tuna, skipjack tuna, hake, Spanish mackerel, little tunny, rainbow trout, sea trout, brook trout);

- tissue samples taken in the laboratory according to a conventional protocol: oral sample from a goat, from a cat; mouse;

- food samples, the exact composition and origin of which are known: blanquette of veal, beef Bourguignon, veal tongue in sauce, joint of lamb, joint of pork, chicken leg;

- commercial samples (denoted "comm" in table 1), obtained from mass marketing, which are beef-based (calves' liver, beefsteak, veal chop, ground beef, joint of veal, Parmentier, Bolognaise), pork-based (ham, sausage, sausages, Chinese pork), poultry-based (ostrich steak, roast chicken, roast guinea-fowl, turkey leg, roast goose) or fish-based (European eel, salted cod fillet, canned yellowtail tuna, canned

skipjack tuna, Atlantic salmon fillet, common mackerel, rainbow trout, arctic char).

All the samples are numbered (E1 to E57), and this numbering was kept in the 5 examples illustrating the invention.

Each sample is placed in a baglight® bag (Intersciences) and then blended until it is homogenized in a BagMixer®-type blender (Intersciences).

b) Lysis of 25 mg of sample and purification of total DNA

The sample is lyzed and nucleic acids are purified using the Dneasy™ tissue kit (Qiagen, ref. 69504), applying the protocol recommended by Qiagen for extracting and purifying the nucleic acids from animal tissues.

c) PCR

A PCR is carried out using the Ampli Taq gold kit from Applied Biosystems according to the protocol below. The following are added to 2 μl of the total DNA suspension: the 10X gold buffer, 3.5 mM of $MgCl_2$, 100 μM of dNTPs (deoxyribonucleoside triphosphates), 2U of Taq gold polymerase, and 0.4 μM of the euvertebrate primers as described by Bartlett et al., in 1992 (Biotechniques Vol. 12 No. 3 pp. 408-412):

SEQ ID No. 233: 5' CCATCCAACA TCTCAGCATG ATGAAA 3' (sequence CDL),

SEQ ID No. 234: 5' **GAAATTAATA CGACTCACTA TAGGGAGACC ACA**CCCCTCA GAATGATATT TGTCCTCA 3' (sequence CBHT7, in bold: T7 polymerase promoter), in order to obtain 50 μl of final reaction volume.

A first PCR cycle of 10 minutes is carried out at 95°C, followed by 35 cycles each made up of the following 3 steps: 94°C for 45 seconds, 50°C for 45 seconds, 72°C for 2 minutes. A final extension of 5 minutes at 72°C is then carried out.

d) Verification of the amplification

In order to verify the amplification, 5 μl of amplification product (or amplicon) are loaded onto a 1.5% agarose gel in an EDTA-Tris borate buffer. After migration for 20 minutes at 100 volts, the amplification band is visualized by staining with ethidium bromide and by illumination with ultraviolet light. The amplification is positive, as demonstrated by the presence of a band having the expected size (350 base pairs).

e) Identification of the amplicon on a DNA chip (Affymetrix, Santa Clara)

A biochip is synthesized on a solid support made of glass according to the method described in US patent 5,744,305 (Affymetrix, Fodor et al.) using the resequencing strategy described in application WO 95/11995 (Affymax, Chee et al.) and according to the method described by A. Troesch et al. (J. Clin. Microbiol., 37(1): 49-55, 1999).

Each identifying sequence comprises 17 bases, with an interrogation position at the 10th position relative to the 3′ end of the sequence.

The analysis is carried out with the GeneChip® complete system (reference 900228, Affymetrix, Santa Clara, CA) which comprises the GeneArray® reader, the GeneChip® fluid station and the GeneChip® analytical software.

e.1. Transcription and labeling of amplicons

Due to the antisense primer CBHT7, all the amplification products have a promoter for T7 RNA polymerase. These amplicons will then serve as a matrix for a transcription reaction during which a fluorescent ribonucleotide will be incorporated.

A 2 µl aliquot is taken from the 50 µl of positive amplification product and is added to a transcription mixture containing the components of the Megascript T7 kit (Ambion, ref. 1334) and fluorescein-12-UTP (Roche, ref. 1427857). The final reaction mixture is prepared in 20 µl and the transcription reaction is carried out for 2 hours at 37°C.

e.2. Fragmentation of the labeled transcripts

In order to improve the hybridization conditions, the labeled transcripts are fragmented into fragments of approximately 20 nucleotides. For this, the 20 µl of labeled transcripts are subjected to the action of 30 mM imidazole (Sigma) and 30 mM manganese chloride (Merck) for 30 minutes at 65°C.

e.3. Hybridization on the DNA chip

A 7 µl aliquot is taken from the 20 µl of labeled and fragmented transcripts and is added to 700 µl of hybridization buffer (6X SSPE (Eurobio)), 5 mM DTAB (Sigma), 3M betaine (Acros), 0.01% antifoam (ref. A80082, Sigma), and 250 µg/ml of herring sperm DNA (Gibco). This mixture is hybridized on the chip under the following conditions: 30 minutes at 40°C. After washing, the chip is scanned and the hybridization image obtained is then analyzed using the GeneChip® software (Affymetrix, Santa Clara, CA).

The hybridization spots make it possible to reconstitute the sequence of the amplicon, which is then compared with the reference sequences of the chip.

The sequence (and therefore the species which corresponds to it) which exhibits the best percentage homology (also called "base-call", expressed as %) with the sequence of the amplicon is selected for the identification.

e.4. Interpretation of the results

Only part of the sequence of 350 bases is analyzed for each species. It corresponds to all or some of the identifying probes. The interpretation threshold, i.e. the level of identification, is set at a 90% base-call on the signature sequence. Below this threshold, the target, and therefore the corresponding species, is not considered to be identified.

f) Result

The DNA extracted from the food sample gives rise to an amplification product, and then to an identification on the chip. As shown in table 1, the reference samples are correctly analyzed by this technique, which also allows the detection of animal species (mammal, bird, fish) in commercial samples.

Table 1: Detection of an animal species in a sample

Animal species	**Nature of the sample**		**% base call Signature sequence**	**Identification on chip**
Cattle (*Bos taurus*)	**ref**	**E1: bovine DNA**	**Bos taurus 100%**	**cattle**
		E2: bourguignon	**Bos taurus 100%**	**cattle**
		E3: veal tongue	**Bos taurus 100%**	**cattle**
		E4: blanquette of veal	**Bos taurus 100%**	**cattle**
	comm	E5: veal chop	Bos taurus 95%	cattle
		E6: ground beef	Bos taurus 100%	cattle
		E7: joint of veal	Bos taurus 100%	cattle
		E8: Parmentier	Bos taurus 100%	cattle
		E9: bolognaise	Bos taurus 100%	cattle

		E10: beef steak	Bos taurus 100%	cattle
		E11: calves' liver	Bos taurus 100%	cattle
Goat (*Capra hircus*)	**ref**	**E12: goat DNA**	**Capra hircus 100%**	**goat**
		E13: oral sample	**Capra hircus 100%**	**goat**
Sheep (*Ovis aries*)	**ref**	**E14: sheep DNA**	**Ovis aries 95.5%**	**sheep**
		E15: joint of lamb	**Ovis aries 100%**	**sheep**
Pig (*Sus scrofa*)	**ref**	**E16: pig DNA**	**Sus scrofa 100%**	**pig**
		E17: joint of pork	**Sus scrofa 100%**	**pig**
	comm	E18: ham	Sus scrofa 100%	pig
		E19: sausage	Sus scrofa 100%	pig
		E20: sausages	Sus scrofa 100%	pig
		E21: Chinese pork	Sus scrofa 100%	pig
Rabbit (*Oryctolagus cuniculus*)	**ref**	**E22: rabbit DNA**	**Oryctolagus cuniculus 100%**	**rabbit**
Hare (*Lepus cuniculus*)	**ref**	**E22: hare DNA**	**Lepus cuniculus 100%**	**hare**
Reindeer (*Rangifer tarandus*)	**ref**	**E23: reindeer DNA**	**Rangifer tarandus 100%**	**reindeer**
Mouse (*Mus musculus*)	**ref**	**E24: mouse**	**Mus musculus 100%**	**mouse**
Cat (*Felis catus*)	**ref**	**E25: oral sample**	**Felis catus 100%**	**cat**
Ostrich (*Struthio camelus*)	**ref**	**E26: ostrich DNA**	**Struthio camelus 100%**	**ostrich**
	comm	E27: ostrich steak	Struthio camelus 100%	ostrich
Chicken (*Gallus gallus*)	**ref**	**E28: chicken DNA**	**Gallus gallus 100%**	**chicken**
		E29: Chicken leg	**Gallus gallus 94.7%**	**chicken**
	comm	E30: roast chicken	Gallus gallus 100%	chicken
Guinea-fowl (*Numida meleagris*)	comm	E31: roast guinea-fowl	Numida meleagris 100%	guinea-fowl
Turkey (*Meleagris gallopovo*)	**ref**	**E32: turkey DNA**	**Meleagris gallopovo 100%**	**turkey**
		E33: turkey joint	**Meleagris gallopovo 100%**	**turkey**
	comm	E34: turkey	Meleagris	turkey

		legs	gallopovo 100%	
Goose (*Anser anser*)	**ref**	**E35: goose DNA**	**Anser anser 100%**	**goose**
	comm	E36: roast goose	Anser anser 100%	goose
European eel (*Anguilla anguilla*)	comm	E37: whole fish	Anguilla anguilla 100%	European eel
Cod (*Gadus morhua*)	**ref**	**E38: cod DNA**	**Gadus morhua 100%**	**cod**
	comm	E39: salted cod fillet	Gadus morhua 100%	cod
Yellowfin tuna (*Thunnus albacares*)	**ref**	**E40: yellowfin tuna DNA**	**Thunnus 100%**	**tuna**
	comm	E41: canned yellowfin tuna	Thunnus 100%	tuna
Skipjack tuna (*Katsuwonis pelamis*)	**ref**	**E42: skipjack tuna DNA**	**Thunnus 94.7%**	**tuna**
	comm	E43: canned skipjack tuna	Thunnus 94.7%	tuna
Atlantic salmon (*Salmo salar*)	comm	E44: Atlantic salmon fillet	Salmo salar 100%	Atlantic salmon
Hake (*Merluccius merluccius*)	**ref**	**E45: hake DNA**	**Merluccius 94.4%**	**hake**
Spanish mackerel (*Scomber japonicus*)	**ref**	**E46: Spanish mackerel DNA**	**Scomber japonicus 100%**	**Spanish mackerel**
Common mackerel (*Scomber scombrus*)	comm	E47: whole fish	Scomber scombrus 100%	common mackerel
Little tunny (*Euthynnus alleteratus*)	**ref**	**E48: little tunny DNA**	**Euthynnus alleteratus 100%**	**little tunny**
Rainbow trout (*Oncorhyncus mykiss*)	**ref**	**E49: rainbow trout DNA**	**Oncorhyncus mykiss 100%**	**rainbow trout**
	comm	E50: whole fish	Oncorhyncus mykiss 100%	rainbow trout
Sea trout (*Salmo trutta fario*)	**ref**	**E51: sea trout DNA**	**Salmo trutta fario 100%**	**sea trout**
Brook trout (*Salvenius fontinalis*)	**ref**	**E52: brook trout DNA**	**Salvenius fontinalis 100%**	**brook trout**
Arctic char (*Salvenius alpinus*)	comm	E53: whole fish	Salvenius alpinus 100%	Arctic char

Example 2: Detection of several animal species in a sample (table 2)

The experimental conditions concerning the preparation of the samples, the lysis of the samples and the purification of total DNA, the PCR, the verification of the amplification and the identification of the amplicon on a DNA chip (Affymetrix, Santa Clara) are identical to that which is described in example 1.

In this example, several animal species are simultaneously analyzed from the same sample. The analysis is carried out on:

reference samples (denoted "ref", as in example 1) consisting of:

a mixture of DNA originating from two different animal species, in a variable proportion of each of the 2 species,

a mixture of amplicons (obtained according to the protocol of example 1), in a variable proportion of each of the two species;

commercial samples (denoted "comm", as in example 1), derived from mass marketing, comprising several animal species in the same sample.

As presented in table 2, these results show that mixtures of species can be detected simultaneously in the same sample, whether this sample consists of a mixture of DNA, a mixture of amplicons or a commercial sample comprising several species.

Table 2: Detection of several animal species in a sample

Sample	Composition	% base call - signature sequence	Chip identification
1) Mixture of amplicons (after amplification)			
Beef (E1) + turkey (E32)	80% v/v 20% v/v	Bos taurus 100% Meleagris gallopovo 94.1%	cattle and turkey
Beef (E1) + turkey (E32)	50% v/v 50% v/v	Bos taurus 100% Meleagris gallopovo 100%	cattle and turkey
Beef (E1) +	20% v/v	Bos taurus 100%	cattle and

turkey (E32)	80% v/v	Meleagris gallopovo 100%	turkey
2) Mixtures of DNA (before amplification)			
Pork (E16) + rabbit (E22)	50% v/v 50% v/v	Oryctolagus cuniculus 100% Sus scrofa 94.7%	pig and rabbit
Chicken (E22) + turkey (E32)	50% v/v 50% v/v	Gallus gallus 100% Meleagris gallopovo 100%	chicken and turkey
Beef (E1) + turkey (E32)	99.9% v/v 0.1% v/v	Bos taurus 100% Meleagris gallopovo 17.6%	cattle
Beef (E1) + turkey (E32)	99% v/v 1% v/v	Bos taurus 100% Meleagris gallopovo 95.1%	cattle and turkey
Beef (E1) + turkey (E32)	90% v/v 10% v/v	Bos taurus 100% Meleagris gallopovo 100%	cattle and turkey
Beef (E1) + turkey (E32)	50% v/v 50% v/v	Bos taurus 100% Meleagris gallopovo 100%	cattle and turkey
Beef (E1) + turkey (E32)	1% v/v 99% v/v	Bos taurus 100% Meleagris gallopovo 100%	cattle and turkey
Beef (E1) + turkey (E32)	0.1% v/v 99.9% v/v	Bos taurus 91% Meleagris gallopovo 95.1%	turkey
Beef (E1) + mutton (E14)	5% v/v 95% v/v	Bos taurus 96.5% Ovis aries 81.1%	cattle and sheep
Pork (E16) + chicken (E22) + turkey (E32)	33% v/v 33% v/v 33% v/v	Sus scrofa 96.5% Gallus gallus 95.6% Meleagris gallopavo 88.9%	pig, chicken and turkey

3) Commercial products			
Pâté (E54)	pork + poultry	Sus scrofa 100% Meleagris gallopovo 94.1%	pig and turkey
White sausage (E55)	pork + poultry	Sus scrofa 100% Meleagris gallopovo 94.1%	pig and turkey
Kebab burger (E56)	beef + mutton + goat	Bos taurus 100% Capra hircus 94.1% Ovis aries 81.2%	cattle, goat and sheep
Ravioli bolognese (E57)	pork + beef	Sus scrofa 100% Bos taurus 95.8%	cattle and pig
Fromage au saumon [cheese with salmon] (E58)	cows' cheese + salmon	Bos taurus 100% Salmo salar 100%	cattle and salmon
Poultry chipolata (E59)	poultry	Gallus gallus 95% Meleagris gallopavo 88%	turkey and chicken
Torti and fricadelles (E60)	pork + poultry	Sus scrofa 100% Gallus gallus 96.5%	pig and chicken

Example 3: Detection of one or more animal species in meals intended for animal feed

a) Preparation of the sample

The experimental conditions concerning the preparation of the samples are similar to those which are described in example 1. The samples are derived from meals intended for animal feed. These samples (numbered from F1 to F17) were listed beforehand in 4 categories, after analysis of the presence of bone fragments as described by Michard (Revue de l'Alimentation animale [Review of animal feed], vol. 508, pp. 43-48, 1997; reference technique).

A distinction is then made between "negative" samples, when the number of bone fragments is less than 20, "trace" samples when there are more than 20 bone fragments but a proportion of bone present in the sample of less than 0.01%, samples "to be monitored" when the proportion is between 0.01% and 1‰, and the "positive" samples when the proportion is greater than 1‰.

b) Lysis of the sample and purification of total DNA

For lysing the sample and purifying the nucleic acids, the Dneasy™ tissue kit (Qiagen, ref. 69504) is used as described in example 1, along with 25 mg of meal. The technique is adapted in order to eliminate the PCR inhibitors. Specifically, these inhibitors (polyphenols, cations (Ca^{2+}, Fe^{3+}), traces of heavy metals, tannins, carbohydrates, salts (NaCl, nitrites)) are present in plants in considerable amounts and, as a result, in the meals intended for animal feed. This adaptation is as follows:

1- After lysis with the ATL buffer and proteinase K, chelex is added during the DNA purification step (200 µl of InstaGene™ Matrix (BIO-RAD, ref. 732-6030)).

2- After incubation for 30 minutes at 56°C, a centrifugation (5 minutes; 14 000 rpm) is carried out and the extraction is carried out as described in the Qiagen Dneasy™ tissue kit manual.

c) PCR

A PCR is carried out using the Ampli Taq gold kit from Applied Biosystems. The following are added to 10 µl of the suspension of meal-extracted total DNA: the 10X gold buffer, 3.5 mM of $MgCl_2$, 100 µM of dNTPs (deoxy-ribonucleoside triphosphates), 2U of Taq gold polymerase, 0.4 µM of the euvertebrate primers CBL and CBHT7 as defined in example 1, in order to obtain 50 µl of final reaction volume. A first PCR cycle of 10 minutes at 95°C is performed, followed by 35 cycles each composed of the following 3 steps: 94°C 45 sec, 50°C 45 sec, 72°C 2 minutes. A final extension of 5 minutes at 72°C is then performed.

d) Verification of the amplification

The amplification is verified as described in example 1.

e) Identification of the amplicon on a DNA chip (Affymetrix, Santa Clara).

This identification step is carried out as described in example 1.

f) Result

The results obtained are given in table 3, and compared with the results obtained by means of the conventional protocol of the prior art. There is complete agreement between the 2 techniques, but with, in addition, indication of the species in the case of the invention. The invention makes it possible to detect the presence

of one or more animal species in samples of meals intended for animal feed.

Table 3: Detection of one or more animal species in meals intended for animal feed

	Conventional protocol		Protocol according to the invention
	Category	Bone fragments	
F1	Negative	< 20 fragments	No species detected
F2	Negative	< 20 fragments	No species detected
F3	Negative	< 20 fragments	No species detected
F4	Negative	< 20 fragments	No species detected
F5	Trace	< 0.01%	No species detected
F6	Trace	< 0.01%	No species detected
F7	Trace	< 0.01%	Pig
F8	Trace	< 0.01%	No species detected
F9	Trace	< 0.01%	Pig, mouse, cattle
F10	To be monitored	0.05%	Pig, cattle
F11	To be monitored	0.03%	Pig, cattle
F12	To be monitored	0.02%	Pig, rat, cattle
F13	To be monitored	0.01%	Pig
F14	Positive	0.23%	Pig, cattle
F15	Positive	0.23%	Cattle, pig
F16	Positive	4.70%	Cattle, pig, mouse, turkey
F17	Positive	3.50%	Cattle, mouse, pig, chicken

Example 4: Detection of the class of the species contained in a sample (table 4)

The aim of this example is to obtain a technique for detecting the vertebrate class (mammals, birds, fish, etc.) of the original animal of the ingredient contained in a food sample or a sample of meal intended for animal feed.

The experimental conditions concerning a) the preparation of the sample, b) the lysis of the sample and the purification of total DNA, c) the PCR, d) the verification of the amplification and e) the identification of the amplicon on a DNA chip

(Affymetrix, Santa Clara), are similar to that which is described in examples 1 and 3.

Identification of the presence of a mammal and/or fish and/or birds is determined by the presence of signatures specific for each class.

For example, for detecting the presence of mammals, use will be made of the signature sequence M1, corresponding to the sequence SEQ ID No. 235 GACA**CAA**CAA CAGC, positions 14685 to 14698 (genbank *Bos taurus* reference sequence; accession No. V00654). The CAA bases at positions 14689-14690-14691 (genbank *Bos taurus* reference sequence; accession No. V00654) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of these three bases at the positions indicated above thus makes it possible to determine the presence of mammals in the sample.

Identification of the presence of birds is determined by the signatures:

O1, corresponding to the sequence SEQ ID No. 236 TCC**CT**AGCCT TCTC, positions 15073 to 15086 (*Gallus gallus* reference sequence; genbank accession No. X52392). The **CT** bases (positions 15076-15077) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of these two bases at the positions indicated above thus makes it possible

to determine the presence of birds in the sample.

O2, corresponding to the sequence SEQ ID No. 237 ACA**CT**TGCCG GAAC, positions 15098 to 15111 (*Gallus gallus* reference sequence; genbank accession No. X52392). The **CT** or **CA** bases (positions 15101-15102) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of these two bases at the positions indicated above thus makes it possible to determine the presence of birds in the sample.

Identification of the presence of mammals and of birds is determined by means of the signature V1, corresponding to the sequence SEQ ID No. 238 ATA**GC**CACAGCATT, positions 14883 to 14896 (genbank *Bos taurus* reference sequence; accession No. V00654). The **GC** bases (at positions 14886 and 14887) are conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds and mammals. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds and mammals. The presence of these two bases at the positions indicated above thus makes it possible to determine the presence of mammals and of birds in the sample.

Identification of the presence of fish is determined by the signature P1, corresponding to the sequence SEQ ID No. 239 ATA**ATA**ACCTCTTT, positions 14713 to 14726 (*Gadus morhua* reference sequence; genbank accession No. X99772). The **ATA** or **ATG** bases (positions 14716-14717-14718) are conserved for all the nucleic acid material corresponding to the predefined species making up the

group that it is desired to investigate, in this case fish. No more than 4 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of these three bases at the positions indicated above thus makes it possible to determine the presence of fish in the sample.

As shown in table 4, this technique makes it possible to detect the presence of mammals and/or birds and/or fish, whether these species are present on their own or as a mixture.

Table 4a: Detection of the class of species in a sample

Samples	Signatures detected	Interpretation
E1: bovine DNA	V1 and M1	mammal
E16: pig DNA	V1 and M1	mammal
E17: joint of pork	V1 and M1	mammal
E12: goat DNA	V1 and M1	mammal
E13: oral sample from goat	V1 and M1	mammal
E35: goose DNA	V1 and O1 and O2	bird
E49: rainbow trout DNA	P1	fish
E51: sea trout DNA	P1	fish
Bovine/turkey amplicon mixture	V1 and M1 and O1 and O2	mammal/bird
E15: joint of lamb	V1 and M1	mammal
F9: "trace" meal	V1 and M1	mammal
F1: "negative" meal	*No positive signatures*	*no identification*
Meal	P1	fish

A variant consists in selecting not a triplet of nucleotides, but a single nucleotide representative of a given class of species.

For example, for detecting the presence of mammals, use

will be made, without distinction, of:

1/ The signature sequence M2, corresponding to the sequence SEQ ID No. 262 CTAATCCTACAAATC, positions 14634 to 14648 (genbank *Bos taurus* reference sequence; accession No. V00654). The T base at position 14641 (genbank *Bos taurus* reference sequence; accession No. V00654) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of this base at the position indicated above thus makes it possible to determine the presence of mammals in the sample.

2/ The signature sequence M3, corresponding to the sequence SEQ ID No. 263 AGCTTCAATGTTTTT, positions 14771 to 14785 (genbank *Bos taurus* reference sequence; accession No. V00654). The A base at position 14778 (genbank *Bos taurus* reference sequence; accession No. V00654) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case mammals. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen mammals. The presence of this base at the position indicated above thus makes it possible to determine the presence of mammals in the sample.

For detecting birds, use may be made, without distinction, of:

1/ The signature sequence O3, corresponding to the sequence SEQ ID No. 264 CGGCCTACTACTAGC, positions 15036 to 15050 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position

15043 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

2/ The signature sequence O4, corresponding to the sequence SEQ ID No. 265 CACATCCCTAGCCTT, positions 15069 to 15083 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15076 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

3/ The signature sequence O5, corresponding to the sequence SEQ ID No. 266 GCCCACACTTGCCGG, positions 15094 to 15108 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15101 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid

material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

4/ The signature sequence O6, corresponding to the sequence SEQ ID No. 267 TTGCCGGAACGTACA, positions 15102 to 15116 (genbank *Gallus gallus* reference sequence; accession No. X52392). The A base at position 15109 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

5/ The signature sequence O7, corresponding to the sequence SEQ ID No. 268 GAACGTACAATACGG, positions 15108 to 15122 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15115 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

6/ The signature sequence O8, corresponding to the sequence SEQ ID No. 269 TGAAACACAGGAGTA, positions

15232 to 15246 (genbank *Gallus gallus* reference sequence; accession No. X52392). The C base at position 15239 (genbank *Gallus gallus* reference sequence; accession No. X52392) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case birds. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen birds. The presence of this base at the position indicated above thus makes it possible to determine the presence of birds in the sample.

For detecting fish, use may be made, without distinction, of:

1/ The signature sequence P2, corresponding to the sequence SEQ ID No. 270 TCAGACATCGAGACA, positions 14512 to 14526 (genbank *Gadus morhua* reference sequence; accession No. X99772). The T base at position 14519 (genbank *Gadus morhua* reference sequence; accession No. X99772) is conserved for all the nucleic acid material corresponding to the predefined species making up the group that it is desired to investigate, in this case fish. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of this base at the position indicated above thus makes it possible to determine the presence of fish in the sample.

2/ The signature sequence P3, corresponding to the sequence SEQ ID No. 271 GTAATAATAACCTCT, positions 14710 to 14724 (genbank *Gadus morhua* reference sequence; accession No. X99772). The T base at position 14717 (genbank *Gadus morhua* reference sequence; accession No. X99772) is conserved for all the nucleic acid material corresponding to the predefined species

making up the group that it is desired to investigate, in this case fish. No more than 5 mutated positions are observed for the remainder of the cited signature for all the sequences making up the nucleic acid material of the group of chosen fish. The presence of this base at the position indicated above thus makes it possible to determine the presence of fish in the sample.

As shown in table 4b, this technique makes it possible to detect the presence of mammals and/or of birds and/or of fish in a sample, in particular a food sample.

Table 4b: Detection of a class of species in a sample

Samples	Signatures detected	Interpretation
Pork liver pâté	M3	Mammals
Beef	M4	Mammals
Chicken	O3 and O4 and O5 and O6	Birds
Chicken paella	O3 and O4 and O5 and O6 and O7	Birds
Spanish mackerel	P2	Fish
Canned sardine	P3	Fish
Fish meal	P2	Fish
Fish meal	P3	Fish
Fresh guinea-fowl	O1, O2, O3, O4, O5, O6, O7 and O8	Birds

Example 5: Universal primers for vertebrate amplification (table 5a and 5b)

The aim of the experiments presented in this example is to obtain primers which are even more sensitive than those described in the preceding examples, and more universal for detecting species in mixtures. In fact, the primers used in examples 1 to 4 are very sensitive with respect to bovine species, which can sometimes mask the presence of other species when they are present in a very small proportion.

Several pairs of primers were used in this example:

A first pair of primers comprising the following sequences

SEQ ID No. 240: 5′ GACCTCCCAG CCCCATCAAA 3′ (sequence CBL 20) and

SEQ ID No. 241: 5′ **GAAATTAATA CGACTCACTA TAGGGAGACC ACA**CAGAATG ATATTTGTCC TCA 3′ (sequence CBHT7 20, with, in bold, the location of the T7 polymerase promoter) was chosen, initially, to increase the threshold of detection of certain species, in particular turkey or sheep, which, when they are in trace amounts in a commercial sample, can be masked by the presence of bovine species.

The technique used to obtain the identification on the chip is as described in example 1a, 1b, 1c (with the modified primers), 1d, 1e.

As shown in table 5a, the use of these new primers makes it possible to obtain, in turkey, a threshold of detection of the order of 1% compared with the primers of examples 1 to 4 where the threshold of detection was of the order of 10%. The use of these new primers also makes it possible, in commercial samples originating from mass marketing, to identify animal species, in particular sheep species, present in trace amounts, which were masked by the presence of bovine species in the preceding examples (table 5b).

Table 5a: Threshold of detection of turkey species in a mixture with bovine species

		Detection on chip: % base call			
% DNA		Primers ex. 1 to 4		Primers ex. 5	
E1: bovine	E32: turkey	bovine	turkey	bovine	turkey
100	0	100	5.9	100	29.4

99.9	0.1	100	17.6	100	41.2
99	1	100	76.5	100	94.1
90	10	100	100	100	100
50	50	100	100	100	100
1	99	100	100	90	100
0.1	99.9	100	100	60	100
0	100	50	94.1	26.9	100
Threshold of detection		0.10%	10%	1%	1%

Table 5b: Detection of sheep species in a mixture with other species

Commercial products	Composition indicated	Detection on chip: species detected	
		Primers ex. 1 to 4	Primers ex. 5
E56: Kebab burger	Bread, precooked ground meat (mutton, beef), sauce	Bovine	Bovine Sheep
E57: Couscous meatball	Beef, mutton, vegetable material	Bovine	Bovine Sheep

Secondly, a second set of primers was chosen and used in duplex with the pair of primers described in example 1 c: when detecting animal species initially present in canned food, there may be a problem of degradation of the DNA of the animal species that it is desired to detect, in particular in the case of canned fish (for example canned tuna).

The technique used to obtain the identification on the chip is as described in examples 1a, 1b, 1d, 1e, with the exception of step 1c: 2 additional internal primers (in addition to the universal primers), which make it possible to amplify the 350 bp region in two smaller portions, are used. Several pairs of primers are studied, making it possible to amplify the 350 bp region in two regions each of between 114 and 245 bp in length, according to the primers used. Two pairs of primers were then selected for their universal nature.

A first pair of primers (used in duplex 1) comprising the following sequences:
SEQ ID No. 272: 5′ AGAIGCICCGTTTGCGTG 3′ (flanked by the T7 polymerase promoter, and I = inosine)
SEQ ID No. 273: TTCTTCTTTATCTGTITCTA (I = inosine)
was chosen, initially, in order to increase the threshold of detection of certain fish species, in particular when these fish species are present in a can of food.
A second pair of primers (used in duplex 2), comprising the following sequences, was also selected:
SEQ ID No. 274: 5′ RTCICGRCARATGTG 3′ (flanked by the T7 polymerase promoter, and R = A or G, I = inosine)
SEQ ID No. 275: 5′ GTIAAYTWYGGITGACTIATCCG 3′ (M = A or C, R = A or G, Y = C or T, W = A or T, I = inosine).

In a manner comparable to that which is described in example 1c, a PCR is carried out using the Ampli Taq gold kit from Applied Biosystems (4311814). The following are added to 2 μl of the total DNA suspension: the 10X gold buffer, 3.5 mM of $MgCl_2$, 100 μM of dNTPs (deoxyribonucleoside triphosphates), 2U of Taq gold polymerase, 0.2 μM of the universal primers for vertebrates CBL and CBHT7 as presented in example 1c, and 0.2 μM of the primers chosen from the pairs of primers defined above (duplex 1 and duplex 2), in order to obtain 50 μl of final reaction volume. A first PCR cycle of 10 min at 95°C is performed, followed by 35 cycles each composed of the following 3 steps: 94°C 45 sec, 50°C 45 sec, 72°C 2 min. A final extension of 5 min at 72°C is then performed.

The amplification is verified by loading 5 μl of amplification product (amplicon) onto a 1.5% agarose gel in EDTA-Tris borate. After migration for 20 min at 100V, two amplification bands are visualized by staining with ethidium bromide and by UV illumination.

The results obtained using each duplex are shown in

table 5c, and compared with the results obtained by means of a "conventional" amplification using only the universal primers as described in example 1c.

Table 5c: Detection of several fish species in a sample (derived from a can of food)

Sample	% base call - signature sequence		
	Duplex 1	Duplex 2	Simplex according to ex. 1
Canned white tuna (*Thunnus alalunga*)	100%	100%	89.2%
Canned Atlantic salmon (*Salmo salar*)	90%	95%	93%
Canned flaked yellowfin tuna (*Thunnus albacares*)	89.5%	94.7%	No amplification

It appears that the primers used in duplex 1 and 2 give better results and better sensitivity when it is desired to detect the presence of fish, in particular in a can of food.

It is quite evident that each primer can be used with or without the T7 promoter.